Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies*

Eric Ghysels
Department of Economics
University of North Carolina†
and CIRANO

Pedro Santa-Clara
The Anderson School
UCLA‡

Rossen Valkanov
The Anderson School
UCLA§

First Draft: May 2003
This Draft: August 17, 2003

Abstract

We use the MIDAS (Mixed Data Sampling) approach to compare the predictive performance of various models of volatility. The models differ in the specification of current volatility measures (squared returns, absolute returns, realized volatility, realized power, and return ranges), in the use of daily or intra-daily (5-minute) data, and in the length of the past history included in the forecasts. The MIDAS framework allows us to compare models across all these dimensions in a very tightly parameterized fashion. Using high-frequency FX data, we find that daily realized power and daily ranges are surprisingly good predictors of future volatility and outperform models based on realized volatility. Moreover, for all daily models, specifications with about 50 daily lags produce the best forecasts. There are no gains from including longer or shorter lags. Also, generally there are no gains from using high-frequency data. The only exception is when too few lags are included in the model, in which case the predictive performance increases with the use of high-frequency data. In models with high-frequency data, absolute returns are better at forecasting future low-frequency realized volatility than squared returns. We also discuss many issues that are encountered in practice, such as long memory and seasonality.

*We thank Arthur Sinko for able research assistance and Tim Bollerslev, Mike Chernov, Rob Engle, David Hendry, Nour Meddahi, Eric Renault, Neil Shephard, Jonathan Wright as well as seminar participants at Oxford, UNC and USC as well as participants at the Symposium on New Frontiers in Financial Volatility Modelling, Florence, the CIREQ-CIRANO-MITACS conference on Financial Econometrics, Montreal and the Research Triangle Conference, for helpful comments.

†Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, phone: (919) 966-5325, e-mail: eghysels@unc.edu.
‡Los Angeles, CA 90095-1481, phone: (310) 206-6077, e-mail: pedro.santa-clara@anderson.ucla.edu.
§Los Angeles, CA 90095-1481, phone: (310) 825-7246, e-mail: rossen.valkanov@anderson.ucla.edu.

The conditional volatility literature, starting with Engle's (1982) ARCH-class of models, has been successful at capturing the dynamics of return variation using simple parametric models. A measure of that success is the widespread use of such models in all areas of finance by academics and practitioners alike. And while most researchers would agree that it is important to have a good model of conditional volatility, the question of what model to use is still unsettled.

When it comes to forecasting volatility, there are many existing models in addition to the de facto benchmark ARCH/GARCH models of Engle (1982) and Bollerslev (1986) which cast future variance as a polynomial of past squared returns, i.e., $\hat{\sigma}^2_{t+1|t} \equiv A(L) r_t^2$. One alternative is to look for variables, other than squared returns, that relate to future volatility. Ding et al. (1993) and several others show that low-frequency components of volatility might be better captured by absolute returns instead of squared returns. Also, Alizadeh et al. (2002) and Gallant et al. (1999) find daily ranges (high-low price ranges) to be good predictors of volatility. Another rapidly growing research area focuses on data-driven models of realized volatility computed from intra-daily returns sampled at very short intervals such as 5 minutes (Andersen and Bollerslev (1998)).[1] Given the variety of volatility models, it seems natural to ask whether some are clearly dominated by others and whether there are real benefits from using high-frequency data.[2] However, these questions have proven difficult to answer because the considered models are so different in terms of regressors, frequencies, parameterizations, and return histories, that is it difficult to compare them.

We use Mixed Data Sampling (henceforth MIDAS) regression models introduced in Ghysels, Santa-Clara and Valkanov (2002a,b) to provide answers to these questions. MIDAS regressions allow us to run parsimoniously parameterized regressions of data observed at different frequencies. First, we use MIDAS regressions to examine whether future volatility, measured over weekly to monthly horizons, is well predicted by past daily squared returns, absolute daily returns, realized daily volatility, realized daily power (sum of intra-daily absolute returns, a measure proposed by proposed by Barndorff-Nielsen and Shephard (2002c) and Woerner (2002)), and daily range. Since all of the regressors are used within a framework with the same number of parameters and the same number of lags, the results from their forecasts are directly comparable. Moreover, we compare the MIDAS results to

[1] Also, see Andersen et al. (2001, 2002, 2003), Andreou and Ghysels (2002b), Barndorff-Nielsen and Shephard (2001, 2002a,b, 2003), Taylor and Xu (1997), among others.

[2] High-frequency data also suffer from microstructure artifacts, such as bid-ask bounce (Roll, 1984), screen fighting, jumps, and irregular or missing data, all of which can lead to biases in the volatility estimates.

those from traditional volatility models. Hence, the MIDAS setup allows us to determine whether some widely-used regressors of conditional volatility clearly dominate the others. We find that in our foreign exchange dataset, among all daily regressors, the realized power dominates the other variables at all horizons, while the daily range is close behind. The surprising finding that the more theoretically motivated realized volatility is dominated by the realized power and the daily range emphasizes the importance of this exercise. We also find that MIDAS forecasts dominate previously used models. Also, daily lags longer than about 50 days do not help (nor hurt) the forecasts, for any of the regressors.

Second, mixed data regressions allow us to directly project future realized volatility onto high-frequency (say 5-minute) squared and absolute returns without daily pre-filtering and without increasing the number of parameters. Hence, we are able to analyze if there are real benefits from directly using high-frequency data in volatility forecasting. Throughout the paper, we consider a foreign exchange (FX) dataset used in a number of recent papers on predicting volatility with high-frequency data. Also, we focus on predicting future conditional variance from one week to one month horizons, because this is the variance that is widely used for option pricing, portfolio management, and hedging. Surprisingly, we find that there are no real benefits of using high-frequency data, with or without long-memory specifications or seasonal adjustments, when a sufficiently long history of returns is included in the regressions. Forecasts using high-frequency data do not outperform those that use daily regressors with 50 lags or more. More precisely, while high-frequency 5-minute models do slightly better than daily realized volatility models, they do not outperform daily range and daily realized power models. However, there are benefits from using high-frequency data when the prediction model is specified with too few lags (1 to 5 days). It must be noted that none of these results are driven by over-fitting or proliferation of parameters: all MIDAS specifications, daily and 5-minutes, have the same number of estimated parameters.

There are several advantages from using mixed data sampling regressions. They allow us to study, in a unified framework, the forecasting performance of a large class of volatility models which involve: (i) data sampled at different frequencies; (ii) various past data window lengths; and (iii) different regressors. The specification of the regressions combine recent developments regarding estimation of volatility and a not so recent literature on distributed lag models.[3] To summarize the results of the paper, we find that the most important gains

[3]See e.g. Dhrymes (1971) and Sims (1974) for surveys on distributed lag models. Many econometrics textbooks also cover the topic, see e.g. Greene (2000, chap. 17), Judge et al. (1985, chap. 9 - 10), Stock and Watson (2003, chap. 13) Wooldridge (2000, chap. 18), among others.

in forecasting are made by a sufficiently long lag structure parsimoniously parameterized through a MIDAS polynomial specification and a judiciously chosen set of regressors, i.e., realized power or daily range.

The MIDAS regressions can also be used to model asymmetries and the joint forecasting power of the regressors. In fact, Engle and Gallo (2003) use the multiplicative error model (MEM) of Engle (2002) and find improvements in forecasting volatility from the joint use of absolute returns, daily ranges, and realized volatilities using return data from the S&P 500 index. Interestingly enough, their results agree with ours, despite the different dataset and different method, that range-based measures in particular provide a very good forecast of future volatility.

The paper is structured as follows. In a first section we introduce MIDAS volatility models and examine various daily regressors. The second section is devoted to MIDAS regression models involving intra-daily data. In section three, we investigate the robustness of the findings and deal with seasonality and long memory, two topic that deserve special attention. Section four concludes.

1 Daily MIDAS Models of Conditional Volatility

To fix notation, let daily returns be denoted by $r_{t,t-1} = \ln(P_t) - \ln(P_{t-1})$. Throughout the paper the time index t will refer to daily sampling. If the data is sampled at a higher frequency, say m-times in a day, then we will denote the return over this interval as $r_{t,t-1/m}$ $= \ln(P_t) - \ln(P_{t-1/m})$. For instance, in our study, returns are sampled every five minutes from a continuously trading 24 hour financial market (which has 288 five-minute intervals within a trading day), and we will write $r_{t,t-1/288} = \ln(P_t) - \ln(P_{t-1/288})$, which corresponds to the last 5-minute return of day $t-1$.

The objective of interest is predicting the increments in the quadratic variation of the return process over some future period, H, namely we want to predict $Q_{t+H,t}$. Therefore, $H = 1$, means forecasting tomorrow's quadratic variation. We focus on predicting future realized volatility, from one week ($H = 5$) to one month ($H = 20$) horizon. This is the variance that matters for option pricing and portfolio management. Focusing on predicting future increments of quadratic variation allows us to make our analysis directly comparable with a large body of existing literature. The quadratic variation is not observed directly but can be measured with some discretization error. One such measure would be the sum of (future) squared returns, namely $\sum_{j=1}^{Hm}[r_{(t+H)-(j-1)/m,(t+H)-(j-2)/m}]^2$, which we will denote by $\tilde{Q}_{t+H,t}^{(Hm)}$ since it involves a discretization based on Hm intra-daily returns. The superscript in parentheses indicates the number of high-frequency data used to compute the variable.

1.1 The MIDAS Specification

A daily MIDAS volatility model is a regression model:

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k,\theta)\tilde{X}_{t-k,t-k-1}^{(m)} + \varepsilon_{Ht} \tag{1.1}$$

The above specification has three important features when compared to other models of conditional volatility (discussed below). First, the volatility measure on the left-hand side, $\tilde{Q}_{t+H,t}^{(Hm)}$, and the variables on the right-hand side, $\tilde{X}_{t-k,t-k-1}^{(m)}$, might be sampled at different frequencies. Second, the polynomial lag parameters b_H are parameterized to be a function of θ, thereby allowing for a longer history without a proliferation of parameters. Third, the $\tilde{X}_{t-k,t-k-1}^{(m)}$ is one of several measures of past fluctuations in returns. We elaborate on each of these features below.

Sampling at Different Frequencies

In equation (1.1), the volatility is measured at weekly, bi-weekly, tri-weekly, and monthly frequency, whereas the explanatory variables are given over daily intervals. In some cases, the daily right-hand side variables are computed using m high-frequency returns, in which case they are denoted by a superscript (m). However, the explanatory variables need not be observed at daily frequency. More specifically, in the next section, the right-hand side variables are measured at a intra-daily (say 5-minute) frequency. In general, the *MI*xed *DA*ta *S*ampling (or MIDAS) framework allows us to investigate whether forecasts of volatility improve from using high-frequency data and how longer horizon forecasts are affected by high-frequency data. These issues have motivated much of the recent literature on high-frequency data, see Andersen et al. (2001, 2002, 2003), Andreou and Ghysels (2002b), Barndorff-Nielsen and Shephard (2001, 2002a,b, 2003), among others.

Parsimony of Parameterization

Another distinguishing feature of the MIDAS specification is that the lag coefficients $b_H(k,\theta)$ (weights) are not free and unrestricted parameters that need to be estimated. Rather they are a function of θ, where θ is a small-dimensional vector. This reduction of the parameters to estimate is particularly important when one has to estimate a persistent process, such as volatility, where distant shocks in the $\tilde{X}^{(m)}_{t-k,t-k-1}$ are likely to have an impact on current volatility. With daily explanatory variables, the unrestricted specification of the weights would result in a lot of parameters to estimate. The problem would only get worse with higher-frequency data. As we will see below, a suitable parameterization $b_H(k,\theta)$ circumvents the problem of parameter proliferation and of choosing the truncation point k^{max}. Hence, the parameterization $b_H(k,\theta)$ is quite appealing in the MIDAS context.

The weights $b_H(k;\theta)$ are normalized to add up to one, which allows us to estimate a scale parameter ϕ_H. Another restriction that is very useful in the context of volatility models is to insure that the weights $b_H(k;\theta)$ be non-negative. Such a condition guarantees that the volatility process itself is non-negative.[4] In general, there are many ways of parameterizing $b_H(k,\theta)$ to satisfy the additivity and positivity constraints. We focus on one such specification based on the Beta function, which has only two parameters, or $\theta = [\theta_1;\theta_2]$:

$$b_H(k;\theta) = \frac{f(\frac{k}{k^{max}},\theta_1;\theta_2)}{\sum_{j=1}^{k^{max}} f(\frac{j}{k^{max}},\theta_1;\theta_2)} \tag{1.2}$$

where: $f(z,a,b) = z^{a-1}(1-z)^{b-1}/\beta(a,b)$ and $\beta(a,b)$ is based on the Gamma function, namely $\beta(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Specification (1.2) was introduced in Ghysels et al. (2002b) and has several important characteristics. Namely, (1) it provides positive coefficients, which is necessary for *a.s.* positive definiteness of the estimated volatility, (2) with $\theta_1 = 1$ and $\theta_2 > 1$ one has a slowly decaying pattern typical of volatility filters, which means that only one parameter is left to determine the shape, and (3) with $\theta_1 = \theta_2 = 1$ one has equal weights, which corresponds to a rolling estimator of the volatility.[5] Nevertheless, MIDAS regression models are not limited to Beta distributed lag schemes, but for our purpose we

[4] It should parenthetically be noted that Nelson and Cao (1992) note that imposing non-negativity constraints on GARCH parameters is unnecessarily restrictive. They do argue, however, that non-negativity of the Wold representation weights is necessary. Restricting the weights $b_H(k;\theta)$ to be non-negative is very similar to the condition imposed by Nelson and Cao.

[5] The flexibility of the Beta function is well known. It is often used in Bayesian econometrics to impose flexible, yet parsimonious prior distributions. The function can take many shapes, including flat weights, gradually declining weights as well as hump-shaped patterns.

focus our attention one specification, bearing in mind that our findings are robust with respect to alternative specifications such as for example the Exponential Almon lag. We refer to Ghysels et al. (2002a,b) for alternative weight specifications and further details.

To illustrate the issue of parameter proliferation, consider Figure 1. It displays the estimated *unconstrained* parameters of equation (1.1) for lags up to 10 days. The figure contains results from various regressors $\tilde{X}^{(m)}_{t-k,t-k-1}$, such as squared returns, $\tilde{Q}^{(m)}_{t-k,t-k-1}$, as well as absolute daily returns, daily range, and realized power, all of which we discuss at length below. We notice from the results displayed in the figure that the parameter estimates appear to be erratic as the lag increases. Hence, volatility models such as (1.1), whose weights are *not* tightly parameterized, do only well with a small number of lags. It must be noted that the robust performance of ARCH/GARCH models can largely be attributed to capturing the dynamics of a large number of past shocks with only a few parameters. This basic idea is also the insight behind the MIDAS regressions.

Various Regressors

In the MIDAS volatility model (1.1), $\tilde{X}^{(m)}_{t-k,t-k-1}$ can be any variable that has the ability to forecast $\tilde{Q}^{(Hm)}_{t+H,t}$. To put it differently, MIDAS volatility regressions can involve $\tilde{X}^{(m)}_{t-k,t-k-1}$ other than past squared returns or past realized volatility, which are the usual regressors considered in the autoregressive conditional volatility literature. The MIDAS approach puts us in the mind set of regression analysis and prompts us to explore various regressors that have the potential to predict future volatility. A number of regressors, other than past realized volatility or squared returns, have been proposed in various models. Unlike MIDAS, however, these models are typically autoregressive in nature. The MIDAS setup allows us to compare the forecasting ability of different $\tilde{X}^{(m)}_{t-k,t-k-1}$'s and to choose the model with the best forecasting ability. In the context of volatility we consider the following regressors:

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^Q + \phi_H^Q \sum_{k=0}^{k^{max}} b_H^Q(k,\theta)\tilde{Q}_{t-k,t-k-1}^{(m)} + \varepsilon_{Ht}^Q \quad (1.3)$$

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^G + \phi_H^G \sum_{k=0}^{k^{max}} b_H^G(k,\theta)(r_{t-k,t-k-1})^2 + \varepsilon_{Ht}^G \quad (1.4)$$

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^a + \phi_H^a \sum_{k=0}^{k^{max}} b_H^a(k,\theta)|r_{t-k,t-k-1}| + \varepsilon_{Ht}^a \quad (1.5)$$

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^r + \phi_H^r \sum_{k=0}^{k^{max}} b_H^r(k,\theta)[hi-lo]_{t-k,t-k-1} + \varepsilon_{Ht}^r \quad (1.6)$$

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^p + \phi_H^p \sum_{k=0}^{k^{max}} b_H^P(k,\theta)\tilde{P}_{t-k,t-k-1}^{(m)} + \varepsilon_{Ht}^p \quad (1.7)$$

In equation (1.3), past $\tilde{Q}_{t,t-1}^{(m)}$ are used to predict $\tilde{Q}_{t+H,t}^{(Hm)}$. Examples of such models involving *unconstrained* polynomial parameters have been advocated by Andersen et al. (2001, 2002, 2003) and are discussed at length below. Specification (1.4) is a projection of $\tilde{Q}_{t+H,t}^{(Hm)}$ onto lagged daily returns and corresponds to the ARCH/GARCH class of models (under some parameter restrictions).

Equations (1.5) and (1.6) involve projecting $\tilde{Q}_{t+H,t}^{(m)}$ onto past daily absolute returns and daily ranges, respectively, which are two alternative measures of volatility. Therefore they are natural candidate regressors in the MIDAS specification. It is often argued that in the presence of deviations from normality absolute values could be more robust than squared values for conditional variance estimation (see e.g. Davidian and Caroll (1987)) whereas the virtues of daily range have been explored most recently by Alizadeh et al. (2001) and Gallant et al. (1999). Typically, past absolute returns (ranges) are used to predict future absolute returns (ranges). In particular, when absolute returns (daily ranges) are considered, the autoregressive features of absolute returns (daily ranges) are studied and modelled (see e.g. Ding et al. (1993) for absolute returns and Alizadeh et al. (2001) for daily ranges). Hence, the exploration of alternative measures of volatility has been cast in the context of autoregressive schemes. Here we introduce absolute returns and range as alternative regressors and examine whether future realized volatility is well predicted by past absolute returns and ranges. The MIDAS regression format makes this a relatively straightforward exercise.

The last regression (1.7) involves similar arguments, using techniques and developments of more recent date. The preference for absolute returns is a subject that has received much attention recently, see in particular Barndorff-Nielsen and Shephard (2002c) and Woerner (2002). Recall that $\tilde{Q}_{t,t+1}^{(m)}$ is defined as the sum of m intra-daily squared returns. Instead of taking squared returns, say every five minutes, Barndorff-Nielsen and Shephard suggest to consider the sum of high-frequency absolute returns, so called realized power variation $\tilde{P}_{t+1,t}^{(m)}$ is defined as $\sum_{j=1}^{m} |r_{t-(j-1)/m,t-(j-2)/m}|$. Regression (1.7) projects past daily realized power on future realized volatility.

To summarize, the MIDAS framework offers the ability to mix data sampled at different frequency, combined with a tightly parameterized model that allows different regressors to forecast volatility. In the next subsection, we look at the performance of models (1.3)-(1.7).

1.2 Empirical Results with Daily Data

We use the MIDAS specifications above to forecast the volatility in the foreign exchange (FX) market. Our dataset is of 24 hour FX quotes supplied by Olsen and Associates. We use five-minute intra-day DM/US$ returns that cover a ten year period, 1/12/1986 to 30/11/1996. The original sample is 1,052,064 five-minute return observations (2,653 days of 288 five-minute intervals per day). The returns for some days were removed from the sample to avoid having regular and predictable market closures which affect the characterization of the volatility dynamics. A description of how these days are removed is found in Andersen et al. (2001). The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Although the FX market is a 24 hour market, we follow Andersen et al. (2001) and remove the weekends, starting Friday evening GMT until Sunday evening GMT. Therefore a week is actually 5 trading days. From this data, we compute daily realized volatility, daily squared returns, daily absolute returns, daily ranges (high return minus low return), and realized power.[6]

The empirical results from regressions (1.3)-(1.7) are reported in Table 1. The estimation of MIDAS regression models is discussed at length in Ghysels et al. (2002b). In general, the errors of MIDAS regression may feature autocorrelation or heteroskedasticity. However,

[6] We computed the range for the Olsen data set of 24 hour FX quotes. We use five-minute intra-day DM/US$ bid/ask midpoint quotes to compute a daily range. The fact that 5 minute data is used implies there is potentially some measurement error in the range (i.e. it is the 5 minute data range, not the tick-by-tick data range).

in the current application all regressions involve non-overlapping prediction samples to avoid autocorrelation in the residuals due to overlapping prediction horizons. Hence, homoscedasticity is a valid assumption for the errors, adding the assumption that the errors are Gaussian would result in nonlinear least squares (NLS) being asymptotically optimal. Such arguments make estimation rather straightforward to implement and also make hypothesis testing relatively simple. Since we compare many models estimated with different number of parameters we use the adjusted R^2 for comparisons. Arguably, this criterion may be criticized in the context of volatility predictions. In the context of portfolio or risk management the objective function may neither be symmetric nor be directly related to a pure measure of statistical fit. It is common, however, to use the adjusted R^2 as a model comparison yardstick, examples include Andersen and Bollerslev (1998), Andersen et al. (2001a,b, 2003), Andreou and Ghysels (2002b), among many others.

We consider four prediction horizons, from 1 through 4 weeks. The horizons cover most practically relevant cases, particularly in the context of Value-at-Risk computations, option pricing and portfolio management. The regressions are run on a weekly, bi-weekly, tri-weekly, and monthly data sampling schemes with non-overlapping $\tilde{Q}^{(Hm)}_{t+H,t}$, for $H = 5$ days ($1wk$), $H = 10$ ($2wks$), 15 ($3wks$) and 20 ($4wks$) days respectively. The MIDAS regressions are obtained with the Beta lag structure (1.2), and for the daily data we find that imposing $\theta_1 = 1$ as a suitable restriction, not rejected by the data in all cases, resulting in monotonically declining weights.

Table 1 presents the adjusted R^2 from the MIDAS volatility models with various regressors. In the first panel, we use a truncation of 50 lags in the estimation, or $k^{max} = 50$. When we use realized daily variances as regressors (equation (1.3)), the R^2 ranges from 0.311 to 0.405 across the forecasting horizons. This performance is markedly better than that obtained using daily squared returns, equation (1.4), where the goodness of fit is between 0.170 and 0.291. Such results are not surprising, because daily squared returns are a noisy measure of past realized volatility. In fact, this finding parallels the comparison of realized volatility and GARCH models in Andersen et al. (2003). Absolute returns are surprisingly good forecasters of future volatility. The R^2s from regression (1.5) are between 0.237 and 0.348. For all forecasting horizons, they are significantly higher than those obtained with $(r_{t-k,t-k-1})^2$ and in some cases only slightly worse than $\tilde{Q}^{(m)}_{t-k,t-k-1}$.

The fourth forecasting variable, the daily range, is a very good predictor of future volatility. Forecasts with $[hi - lo]_{t-k,t-k-1}$ produce R^2s in the range of 0.331 and 0.414. They are significantly higher than the goodness of fit obtained from $\tilde{Q}^{(m)}_{t-k,t-k-1}$, $(r_{t-k,t-k-1})^2$, and $|r_{t-k,t-k-1}|$ at any horizon. The superior performance of ranges has been noted, see e.g. Gallant et al. (1999), and one advantage of using $[hi - lo]_{t-k,t-k-1}$ is that it arguably does not involve any measurement error issues (see e.g. Alizadeh et al. (2001)), Andersen et al. (2003), Andreou and Ghysels (2002b), Barndorff-Nielsen and Shephard (2001, 2002a,b, 2003a) for further discussion of measurement error issues). It should be noted, however, that the Gallant et al. (1999) approach to range-based volatility prediction shares some features with the regression appearing (1.5). The main difference is that the setup in Gallant et al. (1999) involves a parametric continuous time model and a so called reprojection procedure using simulated samples where parameter proliferation is less of an issue. The use of MIDAS regressions allows us to treat parsimoniously the predictive power of range-based measures in a reduced-form setting without specifying an underlying structural continuous time data generating process for returns.

The last variable, the realized power or $\tilde{P}^{(m)}_{t-k,t-k-1}$, is the most remarkable forecaster of volatility. At various horizons, the R^2s obtained from the realized power are between 0.391 and 0.437, slightly higher than those obtained with $[hi-lo]_{t-k,t-k-1}$, and significantly higher than the other forecasters.

It is clear from the parameterization of the weights that, in the MIDAS framework, increasing the lags from 50 to 250 days does not result in an increase in the number of parameters to estimate. Therefore, the results obtained from 250 daily lags (roughly one year), presented in the second panel of Table 1 are directly comparable to those from 50 lags. Interestingly, for all regressors and all horizons, increasing the number of lags does not improve the precision of the forecasts. However, it does not worsen it either. From several specifications, we found that there are no gains in forecasting power of the FX volatility beyond approximately two to three months of daily lags.

Figures 2 and 3 show the profiles of the estimated MIDAS regression weights for the $1wk$ and $4wks$ horizons, respectively. From the former it appears that the weighting scheme puts almost all of its weight on the first 30 to 50 days. This explains why there are no gains made with increasing the lag length beyond 50 days (It should be noted that the weights appearing in the figures are normalized to sum up to one, and hence have a common scale.) The one week horizon weights in Figure 2 show some interesting differences. The regression using daily squared returns puts the least weight on the most recent observation (which is the *noisy* squared return) and has the slowest decay rate. The fastest decline is featured by the realized power models which put a lot of weight on the first lag. Somewhere in between is the range model. Table 2 complements Figure 2, but the weights are not scaled. Rather they are multiplied by the corresponding ϕ_H. It shows the implied weights and intercept of the different MIDAS regressions. It is interesting to note from Table 2 that MIDAS regressions not involving past returns may result in negative intercepts. There is obviously a scaling issues that get resolved through the regression. The case of range is particularly obvious, as we know that the range in an i.i.d. setting is related by a factor of $4 \ln 2$ or roughly 2.77. When comparing the weights in Tables 2 (50 days) and 3 (250 days), we note that there is not much difference between the two polynomial schemes, as discussed earlier. Tables 2 and 3 not only cover the $1wk$ horizon but also the longer ones up to $4wks$. Figure 3 displays the weights (again scaled to sum up to one for comparison across models) for the four week horizon prediction.

In Table 8 we report for a selection of models the parameter estimates (the table also covers some models that will be discussed later). Because of the scaling of the weights of the polynomial the parameter ϕ_H for the $1wk$ horizon is roughly four times smaller than the one for $4wks$. The parameter θ_2 also increases with the horizon, causing a sharper declining pattern (compare Figures 2 and 3). The parameter θ_2 is also significantly different from one (equal weights are $\theta_1 = \theta_2 =1$). Moreover, although the weights for MIDAS models involving daily lags of $\tilde{Q}^{(m)}_{t,t-1}$ and $\tilde{P}^{(m)}_{t,t-1}$ appear similar we note that the estimates of θ_2 for the latter are higher, again implying more weight being put at longer lags.

To conclude, it appears that realized power and range are the best daily regressors. They outperform by far daily squared and absolute returns as well as past daily realized quadratic variation. From Figure 2, it appears that the weighting schemes look quite similar across regressors though there are significant differences across the parameter estimates. The prediction power, however, appears really to be determined by the judicious choice of regressors once a tightly parameterized polynomial is chosen.

1.3 Comparison with Other Daily Volatility Models

Several models of conditional volatility have been proposed. They can be generalized into the following two expressions:

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^Q + \phi_H^Q \sum_{k=0}^{k^{max}} b_H^Q(k) \tilde{Q}_{t-k,t-k-1}^{(m)} + \varepsilon_{Ht}^Q \tag{1.8}$$

$$\tilde{Q}_{t+H,t}^{(Hm)} = \mu_H^G + \phi_H^G \sum_{k=0}^{k^{max}} b_H^G(k;\theta) (r_{t-k,t-k-1})^2 + \varepsilon_{Ht}^G \tag{1.9}$$

In equation (1.8), past $\tilde{Q}_{t,t-1}^{(m)}$ are used to predict $\tilde{Q}_{t+H,t}^{(m)}$. Such models of the realized quadratic variation, analyzed by Andersen et al. (2001, 2003), Andreou and Ghysels (2002b), Barndorff-Nielsen and Shephard (2001, 2002a,b, 2003a), and Taylor and Xu (1997), often rely on Merton's (1980) arguments that arbitrarily accurate estimates of volatility can be obtained by increasingly finer sampling of returns. The above papers show that the use of high-frequency data is beneficial in predicting volatility. The difference between (1.8) and (1.3) is the specification of the weights b_H^Q. In the realized volatility models, the weights are estimated without imposing a functional form. Such models are estimated with a small number of lags (or k^{max} is small), because longer lags result in more parameters to estimate and the possibility of overfitting. In applications, longer lags often result in lower adjusted R^2s. The MIDAS parameterization does not suffer from this drawback as the same number of parameters is used in the Beta function which approximates the function of the unrestricted weights.

Equation (1.9) is a version of the most widely used specification of conditional volatility, namely, the ARCH-type models of Engle (1982) (see also Bollerslev (1986)). In (1.9), future volatility is projected onto lagged daily squared returns, where the weights are tightly parameterized again via a MIDAS setup.

It is in relation to equations (1.8) and (1.9) that we recognize two advantages of the MIDAS regressions, namely, the parsimony of the lag structure and the flexible choice of regressors. In practice, equation (1.8) is estimated with only a few lags. Clearly, with daily data, taking, say, 5 lags (days) has the appeal that a small number of parameters are estimated. But the question remains whether improvement can be made using a longer history of daily returns of, say, 20, 50 or 250 days? This question is particularly pertinent in the $\tilde{Q}_{t+H,t}^{(Hm)}$ modelling context, since it has been shown that this process is persistent. To put it differently, past shocks are likely to have a long-lasting effect on future return volatility. Hence, should we estimate a process with many more lags? If we do, the parameter proliferation may adversely affect the predictive performance. The result in Figure 1 clearly showed the consequences of adding more lags in an unconstrained setting.

As a final note, it must be noted that intuition behind the performance of (1.8)–that realized volatility estimated from higher-frequency data provides better estimates of the volatility–still holds true. As a quick illustration of that fact, we present in table 4 the forecasting performance of the realized volatility model (1.8), where daily realized volatility is computed from 5-minute and 30-minute data. For comparison purposes, we also display the MIDAS ($r_{t-k,t-k-1}$) results from table 1. As anticipated, the high-frequency 5-minute data provides a much better forecast of $\tilde{Q}_{t+H,t}^{(Hm)}$ at any horizon. Yet, 50 days of past squared returns outperform 1 day of realized volatility, since the former yields adjusted R^2's 10% higher at all horizons. Increasing the lags of daily realized volatility suffices however to outperform any model involving long lags of daily squared returns. While this comparison confirms the finding reported earlier in the literature, they are cast here in the context of MIDAS regressions where the trade-off between lag length and choice of regressor is clearly put in the same framework.

2 MIDAS Regressions and High-Frequency Data

So far we covered two appealing features of MIDAS regressions. First, they provide a parsimonious representation of long lag structures, a feature that is important in models of volatility. Second, breaking away from autoregressive schemes allows us to cast the discussion of model selection not only in terms of lag lengths but also in terms of regressor choice.

Now, we turn attention to another appealing feature of MIDAS volatility models–their ability to mix sampling frequencies. More concretely, we will consider whether high-frequency data results in better forecasts of future volatility. In this section, we revisit the choice of regressors in the context of high-frequency data and discuss information sets, polynomials, and implicit filters. The final subsection further explores the empirical evidence of volatility predictability using high-frequency data.

2.1 Choice of Intra-Daily Regressors

Let us reconsider the argument that arbitrarily accurate estimates of volatility can be obtained by increasingly finer sampling of returns. Unlike equation (1.8), where future increments in quadratic variation are projected onto a short number of lags of daily quadratic variation, we explore a new dimension proper to MIDAS regression. Instead of aggregating intra-daily squared returns or absolute returns to daily measures $\tilde{Q}_{t,t-1}^{(m)}$ and $\tilde{P}_{t,t-1}^{(m)}$ why not run MIDAS regressions directly on the high-frequency data? Using the notation $B(L^{1/m})$ for a distributed lag polynomial in $L^{1/m}$ which has the property $L^{1/m} r_{t,t-1/m} = r_{t-1/m,t-2/m}$, we consider the regressions:

$$\tilde{Q}_{t+H,t}^{(Hm_1)} = \mu_H^s + \phi_H^s B_H^s(L^{1/m_2})[r_{t,t-1/m_2}]^2 + \varepsilon_{Ht}^s \tag{2.1}$$

$$\tilde{Q}_{t+H,t}^{(Hm_1)} = \mu_H^a + \phi_H^a B_H^a(L^{1/m_2})|r_{t,t-1/m_2}| + \varepsilon_{Ht}^a \tag{2.2}$$

where the first equation involves projecting $\tilde{Q}_{t+H,t}^{(Hm_1)}$ onto past squared returns sampled at frequency $1/m_2$. Note that the sampling frequency used to compute quadratic variation (m_1) may differ from the sampling frequency used to compute high frequency returns. The specification appearing in (2.1) is closely related to both the daily realized volatility model in (1.8) and the daily MIDAS regression in (1.1). With, say, 5 days of lags there is a direct comparison between (2.1) and (1.8) since it involves the same information set. Likewise, with 50 days of data, the same comparison holds for (2.1) and (1.1). Similar arguments apply to absolute intra-daily returns used in equation (2.2) and the MIDAS regression involving realized powers as in equation (1.7). We now turn to topics that relate to information sets and polynomials in the next two subsections.

2.2 Information Sets

We will consider first projections involving squared returns, namely:

$$E_L(\tilde{Q}_{t+H,t}^{(Hm_1)}|[r_{t,t-j/m_2}]^2, j \geq 0) \equiv \mu_H^s + \phi_H^s B_H^s(L^{1/m_2})[r_{t,t-1/m_2}]^2 \tag{2.3}$$

where $E_L(.|.)$ denotes a linear projection. When we consider information sets involving past realizations of measured quadratic variation, i.e. $\tilde{Q}^{(m)}_{t-j,t-j-1}$, $j = 1, \ldots$ we obtain:

$$E_L(\tilde{Q}^{(Hm_1)}_{t+H,t}|\tilde{Q}^{(m_1)}_{t-j,t-j-1}, j \geq 0) \equiv \mu^Q_H + \phi^Q_H B_H(L)\tilde{Q}^{(m_1)}_{t,t-1}. \tag{2.4}$$

Since the information set $[\tilde{Q}^{(m)}_{t-j,t-j-1}, j \geq 0]$ is a subset of $[r_{t,t-j/m}]^2, j \geq 0$ (assuming a common m) we know the following property holds for linear projections and quadratic loss functions:

$$MSE(E_L(\tilde{Q}^{(m)}_{t+H,t}|[r_{t,t-j/m}]^2, j \geq 0)) \leq MSE(E_L(\tilde{Q}^{(m)}_{t+H,t}|\tilde{Q}^{(m)}_{t-j,t-j-1}, j \geq 0)) \tag{2.5}$$

where MSE stands for the Mean Squared Error. Other criteria, like R^2s would have similar features. The reduction of information from $[r_{t,t-j/m}]^2, j \geq 0$ to $[\tilde{Q}^{(m)}_{t-j,t-j-1}, j \geq 0]$ has the resulting (linear) projection only in terms of daily lags, whereas the linear projection (2.1) involves many lags of high frequency data. On the one hand, the relation in (2.5) may not be attained in practice as the linear projection in (2.3) involves a very large number of parameters. On the other hand, a parsimonious model built on a parametric specification of the high-frequency dynamics may yield very complicated projection formulas. Mixed data sampling regressions exploit a much larger information set, without the cost of parameter proliferation, which is the essential feature defining mixed data sampling regression models. Hence, we try the keep the rich information set without being penalized by prohibitively large parameter spaces.

It should also be noted that the MIDAS regression approach is explicit about the information sets involved. Namely, if one considers two sampling frequencies m_1 and m_2 with, for instance, $m_1 < m_2$, then there is a MIDAS regression model associated with each one of the sampling frequencies. Such MIDAS regressions, involving linear projections $E_L(\tilde{Q}^{(Hm_1)}_{t+H,t}|[r_{t,t-j/m_1}]^2, j \geq 0)$ and $E_L(\tilde{Q}^{(Hm_1)}_{t+H,t}|[r_{t,t-j/m_2}]^2, j \geq 0)$, could be viewed as nested models, though this may not lead to any practical testable restrictions. However, we can use standard regression methods to appraise the fit of the two projections. Namely, in practice we may run into so-called microstructure noise (see for instance Andreou and Ghysels (2002a) for further discussion) and linear regressions with m too high may therefore under-perform. The MIDAS regression framework and the explicit treatment of information sets allows us to address these issues rather directly, instead of assessing separately first the quality of realized volatility extractions and then proceed in a second stage with model specification.

The arguments regarding information sets also apply to absolute returns and realized power specifications. Consider the following linear projection:

$$E_L(\tilde{Q}_{t+H,t}^{(Hm_1)}||r_{t,t-j/m_2}|, j \geq 0) \equiv \mu_H^a + \phi_H^a B_H^a(L^{1/m_2})|r_{t,t-1/m_2}| \tag{2.6}$$

and the linear projection onto past realized absolute power defined as $\tilde{P}_{t-j,t-j-1}^{(m)}$, yielding:

$$E_L(\tilde{Q}_{t+H,t}^{(Hm_1)}|\tilde{P}_{t-j,t-j-1}^{(m_1)}, j \geq 0) \equiv \mu_H^P + \phi_H^P B_H^P(L)\tilde{P}_{t,t-1}^{(m_1)} \tag{2.7}$$

The arguments regarding information sets apply to these two projections as well.[7]

2.3 Polynomials and Implicit Filters

The daily regression (1.8) involving quadratic variation can be viewed as factorizing the polynomial $B(L^{1/m})$ in equation (2.1) into subpolynomials that are directly interpretable. For example, in the linear projection (2.3) one could consider a factorization of the following type: $B(L^{1/m}) \equiv B_d(L)B_{rv}(L^{1/m})$, where $B_d(L)$ is a polynomial in lag L (hence involving daily sampling) and $B_{rv}(L^{1/m})$ is a predetermined polynomial representing realized volatility, i.e. $B_{rv}(L^{1/m}) \equiv \sum_{j=1}^{m}(L^{1/m})^j$ which results in taking the sum of $1/m$ sampled squared returns and hence computing $Q_{t,t-1}^{(m)}$. As noted before, and discussed at length notably by Andreou and Ghysels (2002b) and Barndorff-Nielsen and Shephard (2001, 2002a,b, 2003), realized volatilities are estimates of the quadratic variation of a diffusion subject to measurement error. The presence of measurement error has prompted Andreou and Ghysels (2002b) to consider rolling sample estimators of quadratic variation, called HQV, referring to historical quadratic variation. MIDAS regression volatility models also include factorizations involving HQV estimators. For example, one could consider a factorization of the following type: $B(L^{1/m}) \equiv B_d(L)[B_w(L)B_{rv}(L^{1/m})]$, where $B_w(L)$ represents a rolling sample predetermined polynomial combining daily realized volatilities. Alternatively, one could also consider the intra-daily smoothing proposed by Andreou and Ghysels (2002b), which consists of taking $B_d(L)[B_w(L^{1/m})B_{rv}(L^{1/m})]$, with weighting schemes rolling through intra-daily quadratic variation estimators. It may be clear by now that the direct estimation

[7] It should be noted however, that there are differences in the asymptotics, namely when $m \to \infty$. With $\tilde{P}$ we need a $\sqrt{m}$ normalization, whereas with $\tilde{Q}$ there is no need to normalize by m as shown by Barndorff-Nielsen and Shephard (2002c). We do not endeavor into the asymptotics of $m \to \infty$ in the context of this paper.

of MIDAS regression volatility models, is a simple approach to accomplish two operations in one single step. The first operation is filtering past high-frequency returns such that they represent past daily volatilities. The second operation is the projection of the past filtered volatilities onto future realizations to accomplish optimal predictions. The combination of the two operations into one step has many advantages. For instance, with a flexible enough lag structure, one can accomplish optimal forecasting that exploits more directly the sigma algebra generated by the high frequency data set. Combining high-frequency data first into daily estimates entails a loss of information that cannot be recovered in the second step. Moreover, the design of extraction filters, i.e., is the choice of quadratic variation estimator, is by-passed and therefore the worry about noisy volatility estimates is avoided.

2.4 Empirical Results with High-Frequency Data

With 5 days of lagged realized volatility we obtained R^2's ranging from 0.337 ($1wk$) to 0.386 ($4wks$), results that are reproduced for convenience in the top panel of Table 5. Consider now a MIDAS regression with 5 days of high-frequency squared returns. We note from the results in Table 5 that using high-frequency data typically has a negligible impact on the forecasting performance, namely the adjusted R^2's now range from 0.331 to 0.391 for the 1 and 4 week horizons. Note however, that the unconstrained model with 1 day tells a different story. If only one day is used, then using high-frequency data has a substantial impact, as the adjusted R^2 now increases from 0.275 to 0.323 for the $4wks$ prediction horizon. When we examine the comparison between intra-daily absolute and squared returns we note from Table 5 there is again a dramatic effect of using high-frequency absolute returns rather than squared returns. The adjusted R^2's now is as high as 0.437 compared to 0.391 for squared returns and a four week horizon. Bollerslev and Wright (2001) report results that are suggestive of our finding. In particular, they formulate a model to predict future $\tilde{P}_{t+H,t}^{(Hm)}$ rather than $\tilde{Q}_{t+H,t}^{(Hm)}$, using past high-frequency absolute returns. The resulting predictions are then re-scaled via an "ad hoc" ratio of sample means of $\tilde{Q}_{t+H,t}^{(Hm)}$ and $\tilde{P}_{t+H,t}^{(Hm)}$. They find that this procedure dominates the ones involving past squared returns. The MIDAS regression framework allows use to handle this question as a choice of regressors. The answer to the question which high frequency return series should be used to predict volatility is fairly clear, absolute returns should be preferred to squared returns. Note, however that the adjusted R^2's of .437 attained with high-frequency data is the *same* as that obtained with daily realized power, reported in Table 1. Hence, there appears to be no gain of using the intra-daily information. However, though not reported in the table, when the number of days is again too short (like one day) for realized power, then the use of intra-daily data pays off. This result lines up with the use of a single day of quadratic variation and intra-daily squared returns. Comparing results involving 250 days with those involving 50 days tell us again that there is no need to fit longer lags, as previously reported.

We now turn to the regressions involving 50 lags. Comparing the daily power variation model with the MIDAS regression using high frequency data we find out from Table 5 that there some gains to be made at $1wk$ up to $3wks$ as the MIDAS regression fits better. At the $4wks$ horizon we no gain in terms of adjusted R^2's as the polynomial does not seem to fit as well the longer series resulting in a slight reduction of fit. With squared returns there are gains from the 5 day to the 50 specification at all horizons, however.

It should be noted that all the MIDAS regression models using intra-daily data did *not* involve the restriction $\theta_1 = 1$, and in fact this restriction, implying a downward sloping weighting scheme is strongly rejected. More can be learned by looking at plots to understand this finding. In Figure 4 and 5 we plot the polynomials for 5 days of high frequency squared and absolute returns for the one and four weeks horizons. At the $1wk$ horizon we see a sharp decline in the weights. In contrast, at the longer $4wk$ horizon both models now show a peak at the second or third day. In comparison to the daily weights in Figures 2 and 3 the weights changed from slowly decaying to a hump-shaped pattern. Such results are consistent with a recent literature on multi-factor variance models, see e.g. Alizadeh et al. (2002), Chernov et al. (2003), Engle and Lee (1999), among others. The findings show a first factor with high persistence and low volatility, whereas the second factor is transitory and highly volatile. As one predicts longer horizons, the short term factor becomes less important and implies a down-weighing of the most recent observations. While for daily data we did not recover such hump-shaped patterns (which lead us to specify $\theta_1 = 1$), it appears that high-frequency returns contain too much of a noise component that needs to be filtered out via the hump-shaped pattern of MIDAS weights. The flexibility of MIDAS regressions allowed us to accommodate this. In the next section we will address some of this noise issues through seasonal adjustment.

One final issue to note in Table 5 is the comparison between 30 minute MIDAS regressions and 5 minute ones. Andersen et al. (2002, 2003) have used both and appear to use the latter more recently, arguing that microstructure noise may affect the reliability of measures of quadratic variation. We noted before that the MIDAS regression approach is explicit about the information sets involved. In particular, MIDAS regressions, involving linear projections with $m_1 < m_2$ sampling frequencies could be viewed as nested models. Using standard regression methods we can appraise the fit of the two projections and overall we find that the five minute data yields almost uniformly better models. This does not mean that the data is not affected by microstructure noise. It only means that the regressions using five minute data appear to do well. Questions whether there is microstructure noise at those frequencies need to be addressed by other techniques (see for instance Andreou and Ghysels (2002a) for further discussion). The MIDAS regressions only help us to appraise how far we can go in tackling higher frequency data. In the empirical example we provide, going from 30 minutes to 5 is beneficial.

Table 9 displays parameter estimates for a selection of models involving intra-daily data (including again models not considered so far). The relation of one-to-four is again apparent in the scale of ϕ_H between $1wk$ and $4wks$ horizons. More importantly, the parameter θ_1, which were fixed in the daily data models, is now estimated and all values are significantly different from 1. Yet, at the $1wk$ horizon, consistent with Figure 4, we note parameter estimates close to 1. In contrast for the longer horizon we note values around 1.5 which yield the hump shapes appearing in Figure 5. In comparison to the daily models θ_2 takes on smaller values.

3 Robustness of Findings

In this section we examine whether the empirical findings reported so far are robust to at least two issues that we ignored so far. These issues are (1) the presence of seasonality in high-frequency financial returns and (2) the possibility of long memory modelling through fractional integration. A subsection is devoted to each of the issues.

3.1 Dealing with Seasonality

Intra-daily seasonal patterns in financial markets have been widely documented in the literature since the work of Wood et al. (1985).[8] Clearly tomorrow's quadratic variation

[8] The stylized example of FX financial markets operating on a twenty four hour basis features fluctuations of activity that depend on the regional opening and closing of markets around the globe, starting with Asian markets early in the day and tapering off with North American ones.

does not feature seasonal intra-daily fluctuations, whereas the high-frequency returns do. The most practical way to address such a situation is to remove the high frequency intra-daily seasonal patterns from returns prior to fitting the lag structures. One may therefore wonder whether one could improve MIDAS regressions involving intra-daily data by considering seasonal adjustment prior to fitting a MIDAS polynomial. One could accommodate the intra-daily patterns directly via the specification of the polynomial lag structure. A mixture of Beta lag polynomials can accommodate this by centering one polynomial on the mean of a seasonal lag. For example, suppose one day's worth of lags is 288, then there are 24 hourly peaks in the lag structure. Therefore one can take $\theta = ((a_i, b_i), i = 1, \ldots, 24)$, and center the Beta polynomials at the hourly lags, i.e. $a_i/(a_i + b_i) = 12i$ for $i = 1, \ldots, 24$. This would not be very appealing, however, as one would involve at least 48 parameters, to obtain a polynomial lag structure flexible enough to accommodate all the ups and downs of periodic market activity throughout a business day. Note that in less complex seasonal settings, involving quarterly or monthly data, the strategy may indeed be appealing. The alternative is to seasonally adjust the data prior to fitting MIDAS regressions, a subject to which we turn next.

3.1.1 Seasonality in MIDAS Regressions

There are theoretical arguments that support seasonally adjusting the data prior to fitting a MIDAS regression model, and those arguments are spelled out in detail in Hansen and Sargent (1993) and Sims (1993). We consider again a generic regressor $x^{(m)}$, which may be squared returns, absolute returns or range, that is adjusted for (intra-daily) seasonal patterns. We denote by $x^{(m)}$ the raw regressors, and $\tilde{x}^{(m)}$ the adjusted ones. We have considered two alternative schemes for removing the periodic patterns. The first consists of subtracting hourly means of squared returns, i.e. $[\tilde{x}^{(m)}_{t-j/m}]^2 \equiv [\tilde{x}^{(m)}_{t-j/m}]^2$ - $M_{h(j)}$, where $h(j)$ is the hour to which the j^{th} observation belongs and M_h is the mean of $[x^{(m)}_{t-j/m}]^2$ for hour h. The second scheme is slightly more general as it adds a day-of-the-week effect, namely $[\tilde{x}^{(m)}_{t-j/m}]^2 \equiv [x^{(m)}_{t-j/m}]^2$ - $M_{h(j),d(t)}$, where $d(t)$ is the day of the week of day t and $M_{h,d}$ is the mean of $x^{(m)}_{t-j/m}]^2$ for hour h on day d. Obviously one could consider other adjustment schemes as well.

Let us consider the example of squared return regressors and rewrite (2.1) in terms of adjusted returns as:

$$\tilde{Q}_{t+H,t}^{(Hm_1)} = \mu_H^s + \phi_H^s B_H^s(L^{1/m_2})([\tilde{r}_{t,t-j/m_2}]^2 - M_{h(j),d(t)}) + \varepsilon_{tH}^s$$

which can be rewritten as:

$$\tilde{Q}_{t+H,t}^{(Hm_1)} = \mu_H^s + \phi_H^s B_H^s(L^{1/m_2})[\tilde{r}_{t,t-j/m_2}]^2 + \beta_2 z_t + \varepsilon_{tH}^s \tag{3.1}$$

where $\beta_2 z_t$ represents a set of seasonal dummies, i.e. $z_t \equiv (z_t^d, d = 1, \ldots, 7)$ with $z_t^d = 1$ if $d = d(t)$ and zero otherwise. Finally, the parameters β_2 are coefficients implicitly defined by the relationship $B_1(L^{1/m})M_{h(j),d(t)} \equiv \beta_2 z_t$. Comparing equations (3.1) and (2.1) yields the result that in principle one could run a regression with seasonal dummies and raw data to estimate $B_H^s(L^{1/m})$, instead of adjusting squared returns first and estimate the polynomial lag structure with $r_t^{(m)}$. Unfortunately, both are not necessarily equivalent even in large samples. The main reason is that we are dealing with potentially misspecified models since the polynomial lag structure is an approximation to a highly over-parameterized projection. In such circumstances it is often better to adjust series prior to fitting "approximate" models. The formal arguments in Hansen and Sargent (1993) and Sims (1993), are based on the fact that the approximation error may be unduly influenced by the large seasonal variation in the data. In the empirical work involving intra-daily FX data, we will show that adjusting the squared returns prior to running mixed data sampling regressions does not seem to have much of an impact, namely using raw data appears to be doing very well, yielding better results that those obtained with adjusted returns.

3.1.2 Empirical Results

Having compared daily MIDAS with models using high-frequency raw data we now turn to MIDAS models with adjusted returns. We only report a subset of the results we obtained, for example we do not report the hourly/daily adjustment as it yielded essentially results similar to the hourly adjustment scheme. We also refrain from reporting results for all the models considered so far for the same reasons. Table 5 reveals that models involving intra-daily seasonal adjustment under-perform compared to the corresponding model with unadjusted data. In all cases the adjusted R^2s are lower compared to the corresponding unadjusted MIDAS regressions. Note that in none of the cases any explicit penalty was introduced for the use of parameters in estimating the adjustments. The finding probably indicates that more complicated models may be needed to adjust for seasonality in FX returns, see e.g. Andersen and Bollerslev (1987) for further discussion on this matter. However, more complicated adjustment models may also defy the purpose of parsimony. In a different context, the finding may not be true, as there are probably many other settings where seasonals are easier to model. Macroeconomic data series would be a good example to investigate this question further, as a monthly cycle would be easier to capture than a twenty four hour intra-daily setting. The findings reported in Table 5 are significant, however, as one advantage of computing daily measures is that seasonality issues are foregone through the aggregation process. It appears from our findings that the presence of seasonals is not an obstacle to the application of MIDAS models to high-frequency financial data.

3.2 Long Memory

Parsimonious models of volatility often involve fractional differencing, as the phenomenon of long memory has been widely documented. Ding et al. (1993) describe in details slow decay rates of autocorrelation functions of returns and absolute returns, which has given rise to various long-memory GARCH models (Baillie, Bollerslev and Mikkelsen (1996), Bollerlslev and Mikkelsen (1996), Robinson (1991)), and long-memory discrete time stochastic volatility models (Breidt, Crato and De Lima (1998), Harvey (1998)). Comte and Renault (1998) proposed a continuous time stochastic volatility model with long-memory. In this section we consider fractional differencing in the context of MIDAS regressions and report how they perform empirically.

3.2.1 Fractional Differencing and MIDAS Regressions

A time series is said to exhibit long memory if it has a covariance function $\gamma(j)$ that is of the same order as j^{2d-1} (as $j \to \infty$), and a spectrum $f(\lambda)$ that is of the same order as λ^{-2d} (as $\lambda \to 0$) for $0 < 0.5$. Ghysels et al. (2002b) discuss the notion of *reverse engineering* when the MIDAS regression implied by a time series process for the low frequency data is aggregated to yield a MIDAS prediction model. Sinko (2003) reports reverse engineering results for time series processes with long memory. For example, considering an ARFI(1,d) process:

$$(1 - L^{1/m})^d (1 - \phi L^{1/m}) y_{t,t-1/m} = \varepsilon_{t,t-1/m} \tag{3.2}$$

leads to the following MIDAS regression:

$$E(y_{t+1,t}|\varepsilon_{t-j/m,t-(j-1)/m}, \forall j > 0) = \sum_{l=0}^{\infty}(\sum_{j=0}^{l}\beta_j\psi_{l-j}^{(m)})y_{t-l/m,t-(l-1)/m} \tag{3.3}$$

where $\beta_0 = 1, \beta_j = \pi_j\left((-d+j)/(j+1)-\phi\right)$, $\pi_j = \Gamma(-d+j)/(\Gamma(-d)\Gamma(j+1))$, $\psi_k^{(m)} = \sum_{i=0}^{k}\alpha_i^{(m)}\phi^{k-i}$ and

$$\alpha_i^{(m)} = \frac{1}{\Gamma(d+1)}\left[\frac{\Gamma(1+d+m+i)}{\Gamma(1+m+i)} - \frac{\Gamma(d+i+1)}{\Gamma(i+1)}\right].$$

This simple example shows that the reverse engineered MIDAS regression implies a polynomial which is a complex function of the two ARFI parameters d and ϕ. The functional forms of a MIDAS specification will have to mimic the slowly decaying polynomial weights appearing in (3.3), in this case implicit functions of d and ϕ. One can either accomplish this by estimating the polynomial weights directly, as was done so far, or by introducing directly the fractional differencing parameter d in the context of a MIDAS regression. We will refer to such models as FI MIDAS and pursue both approaches in the context of volatility, bearing in mind that at this stage we are not able to say much about the asymptotic distribution theory of MIDAS regression models with and without fractional differencing applied to processes featuring long memory.[9]

The notion of long memory, when applied to volatility involves processes such as $\tilde{Q}_{t,t-1}^{(m)}$, $[r_{t,t-1/m}]^2$, $|r_{t,t-1/m}|$, $\tilde{P}_{t,t-1}^{(m)}$ or $[hi-lo]_{t,t-1}$. In the context of MIDAS regressions we need to highlight a few important properties. First, Andersen and Bollerslev (1997), Harvey (1998), Bollerslev and Wright (2001), among others, show that the log squared returns, squared returns and absolute returns have the same persistent properties. Second, since in MIDAS regressions one mixes data from different sampling frequencies one can exploit another property of fractional differencing, namely its invariance to temporal aggregation. Both properties combined allow us to write:

$$\begin{array}{ccccc}
Cov(\log(r_{t,t-1}^2), \log(r_{t-j,t-j-1}^2)) & \sim & Cov(\log([r_{t,t-1/m}]^2), \log([r_{t-j/m,t-(j-1)/m}]^2)) & \sim & j^{2d-1} \\
Cov(r_{t,t-1}^2, r_{t-j,t-j-1}^2) & \sim & Cov([r_{t,t-1/m}]^2, [r_{t-j/m,t-(j-1)/m}]^2) & \sim & j^{2d-1} \\
Cov(|r_{t,t-1}|, |r_{t-j,t-j-1}|) & \sim & Cov(|r_{t,t-1/m}|, |r_{t-j/m,t-(j-1)/m}|) & \sim & j^{2d-1}
\end{array}$$

[9] It is beyond the scope of the current paper to endeavor into the theory of FI MIDAS models. Sinko (2003) provides some first results on this topic.

where the notation $\sim$ means that the limit of the ratio of the quantities on the left and right hand side is a finite positive constant. Since the empirical measures $\tilde{Q}^{(m)}_{t,t-1}$ and $\tilde{P}^{(m)}_{t,t-1}$ involve sums of $I(d)$ processes over a finite interval we also have the property that:

$$\begin{array}{rcl} Cov(\tilde{Q}^{(m)}_{t,t-1}, \tilde{Q}^{(m)}_{t-j,t-j-1}) & \sim & j^{2d-1} \\ Cov(\tilde{P}^{(m)}_{t,t-1}, \tilde{P}^{(m)}_{t-j,t-j-1}) & \sim & j^{2d-1} \end{array}$$

Exploiting these properties we can first of all note that in a generic MIDAS regression one can model:

$$(1-L)^d y_{t+1,t} = \beta_0 + \beta_1 B(L^{1/m})(1-L^{1/m})^d y_{t,t-1/m} + \varepsilon_t \tag{3.4}$$

with the same fractional differencing operator involved. This feature is rather interesting for the analysis of long memory in MIDAS regressions.[10] Assuming that we can write $B(L^{1/m})(1-L^{1/m})^d$ as $(1-L)^d \tilde{B}(L^{1/m})$ we can also consider the alternative formulation:

$$(1-L)^d y_{t+1,t} = \beta_0 + \beta_1 (1-L)^d \tilde{B}(L^{1/m}) y_{t,t-1/m} + \varepsilon_t \tag{3.5}$$

which applies the fractional differencing operator to the MIDAS polynomial.[11] In the empirical application we will consider cases (3.5) and (3.5), namely we will consider the following long memory specifications:

$$(1-L)^d \tilde{Q}^{(Hm_1)}_{t+H,t} = \mu^s_H + \phi^s_H (1-L)^d B_s(L^{1/m_2})[r_{t-1/m_2}]^2 + \varepsilon^s_{Ht} \tag{3.6}$$

$$(1-L)^d \tilde{Q}^{(Hm_1)}_{t+H,t} = \mu^Q_H + \phi^Q_H (1-L)^d B_Q(L) \tilde{Q}^{(m_1)}_{t,t-1} + \varepsilon^Q_{Ht} \tag{3.7}$$

$$(1-L)^d \tilde{Q}^{(Hm_1)}_{t+H,t} = \mu^a_H + \phi^a_H (1-L)^d B_a(L^{1/m_2})|r_{t,t-1/m_2}| + \varepsilon^a_{Ht} \tag{3.8}$$

$$(1-L)^d \tilde{Q}^{(Hm_1)}_{t+H,t} = \mu^P_H + \phi^P_H \phi^P_H (1-L)^d B_P(L) \tilde{P}^{(m_1)}_{t,t-1} + \varepsilon^P_{Ht} \tag{3.9}$$

Obviously, the estimation theory for FI MIDAS models is, as noted before, not yet established. Ideally, we would like to estimate d and the MIDAS polynomial parameters simultaneously. We proceeded however, in two steps, estimating d first and then proceeding with the estimation of the polynomial parameters next.

[10]This feature does not preclude the fact that $B(L^{1/m})$ in (3.4) still depends on d, as suggested by (3.3), which implies that fully efficient estimation is not accomplished by estimating (3.4) not taking into account this dependence.

[11]The specification in (3.5) is more reminiscent of fractional cointegration specifications when a process $x_{t,t-1/m}$ is involved.

3.2.2 Empirical Results

Models involving fractional differencing will have a prefix FI. For the fractional differencing parameter we used $d = 0.374$ computed with a Geweke-Porter Hudack estimator applied to $\tilde{Q}^{(m)}_{t+H,t}$.

The first question we address is how the different models that capture long memory fare. Comparing the results in Table 4 with those of Table 7 we see for the unconstrained parameter models involving realized volatility up to five days that fractional differencing does indeed pay off at the short horizon, but this advantage evaporates for the longer horizons. For the one day lag model, which de facto is an infinite lag model since it involves fractional differencing, we clearly see the advantage of FI models at all but the four week horizons. Fractional differencing also clearly pays off when daily squared returns are used and also when daily absolute returns are used in the context of MIDAS regressions. The lower part of Table 7 displays results of MIDAS models involving 50 and 250 days of daily or intra-daily data. In none of the cases is there any clear gain from using fractional differencing. Hence, there appears to be no gain of using fractional differencing operators, i.e. the slowly decaying patterns appear to be better fit by the MIDAS regressions polynomials. The flexibility of MIDAS modelling really plays an important factor in these findings. The MIDAS regression weight patterns appearing in Figure 3 clearly show that the monotone declining shape implied by a fractional differencing operator may not be the best way to go. It is perhaps the most telling difference between the long memory GARCH and SV model specifications (as in Baillie, Bollerslev and Mikkelsen (1996), Bollerlslev and Mikkelsen (1996), Breidt, Crato and De Lima (1998), Comte and Renault (1998), Harvey (1998) and Robinson (1991)) and the two factor GARCH and SV models as in Engle and Lee (1999) and Chernov et al. (2003) among others.

Parameter estimates for a selection of FI MIDAS models appear in Tables 8 and 9. There appear to be some problems with the estimation of θ_2 in the case of realized power and $1wk$ horizon, though this is not the case at the four week horizon. One complicating factor that makes comparison rather difficult between FI MIDAS and corresponding MIDAS parameter estimates is that neither the intercept (μ_H) nor the slope (ϕ_H) have the same interpretation, which makes de facto the weighting scheme and implied prediction formula very different. We note that θ_2 is much lower at the short horizon for daily realized volatility. The converse is the case for the longer horizon, though the slope coefficients are very different and they do not apply to the same process. Clearly more work is needed in this area to come to solid conclusions on this matter as the estimation theory of FI MIDAS needs more investigation. We leave this to future research. The evidence presented here is a first brush of paint attempting to get to the whole final picture.

4 Conclusions

We study the predictability of return volatility with MIDAS regressions. Our approach allows us to compare forecasting models with different measures of current volatility, frequencies, and lag lengths. While the main focus of this paper is volatility forecasting, it is clear that the MIDAS framework is general in nature and can find a good use in any empirical investigation that involves data sampled at different frequencies. Simplicity, robustness, and parsimony are three of its main attributes.

We report several surprising findings regarding the predictability of weekly to monthly realized volatility in the FX market. First, we find that daily realized power outperforms daily realized volatility and that daily and intra-daily absolute returns outperform respectively daily and intra-daily squared returns. This set of results suggests that absolute returns are very successful at capturing fluctuations in future return volatility, despite the predominant emphasis in the literature on squared returns. Also, we find that daily ranges are extremely good forecasters of future volatility and are only second to realized power. This last finding is consistent with results in Gallant et al. (1999) and Engle and Gallo (2003), among others, who use different methods and different data.

Interestingly, there are no benefits from directly using high-frequency data. Indeed, intra-daily MIDAS models perform no better with 5-minute returns than with daily measures. The only case when we can report improvements from using high-frequency data is when the lag specification of the model is too short. In a related finding, models that include about 50 lags of daily regressors outperform those with longer or shorter lags. This conclusion is not due to proliferation of lag parameters. In the MIDAS specification, we keep the number of parameters fixed when we consider different lag lengths and different frequencies. This feature of our approach allows us to easily compare models across these dimensions. Moreover, it appears that MIDAS polynomials show the necessary flexibility in capturing the persistence in realized volatility without the need to resort to fractional differencing. In sum, we find that the most important gains in forecasting volatility are made by a sufficiently long past of returns and a judiciously chosen set of regressors.

Finally, we have kept the mixed data sampling regressions as simple as possible in the interest of clarity and conciseness. However, there are a host of issues, such as asymmetries, multiple forecastors, and option-implied measures of volatlity, that merit further attention. These extensions are easily accommodated in the MIDAS framework and some of them are addressed in Ghysels et al. (2002a).

References

[1] Alizadeh, S., M. Brandt and F. X. Diebold, (2002), "Range-based estimation of stochastic volatility models", *Journal of Finance*, 57, 1047-1091.

[2] Andersen, T. and T. Bollerslev (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long Run in high-frequency Returns", *Journal of Finance*, 52, 975-1005.

[3] Andersen, T. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts", *International Economic Review*, 39, 885-905.

[4] Andersen, T., T. Bollerslev, F. X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility", *Journal of American Statistical Association*, 96, 42-55.

[5] Andersen, T., T. Bollerslev, F. X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility", *Econometrica*, (forthcoming).

[6] Andersen, T.G., Bollerslev, T. and F.X. Diebold, (2002), " Parametric and Nonparametric Volatility Measurement," in L.P. Hansen and Y. Aït-Sahalia (eds.), Handbook of Financial Econometrics . Amsterdam: North-Holland, forthcoming.

[7] Andersen, T.G., Bollerslev, T. and N. Meddahi, (2003) "Correcting the Errors: A Note on Volatility Forecast Evaluation Based on High-Frequency Data and Realized Volatilities," Discussion Paper CIRANO.

[8] Andreou, E. and E. Ghysels, (2002a), "When do Jump Diffusions Approximate Returns? Nonparametric Test Procedures Based on Price Rigidity and Discreteness", Work in progress.

[9] Andreou, E. and E. Ghysels (2002b), "Rolling sample volatility estimators: some new theoretical, simulation and empirical results" *Journal of Business and Economic Statistics*, 20, 3, 363-376.

[10] Baillie, R.T., T. Bollerslev and H. O. Mikkelsen (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity" *Journal of Econometrics*, 74, 3-30.

[11] Barndorff-Nielsen, O. and N. Shephard (2001), "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion), *Journal of the Royal Statistical Society*, Series B, 63, 167-241.

[12] Barndorff-Nielsen, O. and N. Shephard (2002a), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models", *Journal of the Royal Statistical Society*, Series B, 64, 253-280.

[13] Barndorff-Nielsen, O. and N. Shephard (2002b) "Estimating quadratic variation using realised variance" *Journal of Applied Econometrics*, 17, 457-477.

[14] Barndorff-Nielsen, O. and N. Shephard (2002c) "Power variation with stochastic volatility and jumps" Discussion paper, Nuffield College.

[15] Barndorff-Nielsen, O. and N. Shephard (2003) "How accurate is the asymptotic approximation to the distribution of realised volatility?" in D.W.K. Andrews, J. Powell, P. Ruud and J. Stock (ed.), *Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg*, Cambridge University Press, (forthcoming).

[16] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

[17] Bollerslev, T., R. Engle and D. Nelson (1994), "ARCH Models, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier Science.

[18] Bollerslev, T. and J. Wright (2001), "High-frequency data, frequency domain inference, and volatility forecasting", *Review of Economics and Statistics*, 83, 596-602.

[19] Breidt, J., N. Cato and P. de Lima (1998) "On the detection and estimation of long memory in stochastic volatility", *Journal of Econometrics*, 83, 325-348.

[20] Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, George (2003), "Alternative Models for Stock Price Dynamics", *Journal of Econometrics*, 116, 225-257.

[21] Comte, F. and E. Renault (1998), "Long Memory Continuous Time Models", *Journal of Econometrics*, 83, 291-323.

[22] Davidian M. and R.J. Carroll (1987), "Variance Function Estimation", *Journal of the American Statistical Association*, 82, 1079-1091.

[23] Dhrymes, P. (1971), *Distributed Lags: Problems of Formulation and Estimation*, Holden-Day, San Francisco.

[24] Ding, Z., and C. W. J. Granger (1996), "Modeling Volatility Persistence of Speculative Returns: A New Approach", *Journal of Econometrics*, 73, 185-215.

[25] Ding, Z., C. W.J Granger and R. F. Engle (1993), "A long memory property of stock market returns and a new models", *Journal of Empirical Finance*, 1, 83-106.

[26] Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1008.

[27] Engle, R.F. and G. Gallo (2003), "A Multiple Indicator Model for Volatility Using Intra Daily Data", Discussion Paper NYU and Universitá di Firenze.

[28] Engle, R.F. and G. Lee (1999), "A Long-Run and Short-Run Component Model of Stock Return Volatility", in R.F. Engle and H. White (eds.) *Cointegration, Causality and Forecasting - A Festschrift in Honour of Clive W. J. Granger*, Oxford University Press.

[29] Fung, W.K.H. and D.A. Hsieh (1991) "Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies" Discussion Paper, Fuqua School of Business, Duke University.

[30] Gallant, A. R., C.-T. Hsu, and Tauchen, G. (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Volatility", *Review of Economics and Statistics*, 84, 617-631.

[31] Ghysels, E., P. Santa-Clara and R. Valkanov (2002a), "There is a Risk-return Tradeoff after all," available at: http://www.personal.anderson.ucla.edu/pedro.santa-clara/.

[32] Ghysels, E., P. Santa-Clara and R. Valkanov (2002b), "The MIDAS touch: Mixed Data Sampling Regression," Discussion paper UCLA and UNC available at: http://www.personal.anderson.ucla.edu/pedro.santa-clara/.

[33] Giraitis, L., P. Kokoszka and R. Leipus (2000), "Stationary ARCH models: dependence structure and Central Limit Theorem", *Econometric Theory*, 16, 3-22, 2000

[34] Granger, C.W.J. (1969), Investigating Causal Relations by Econometric Models and Cross Spectral Methods, *Econometrica*, 37, 428-438.

[35] Greene, W. H. (2000), *Econometic Analysis*, Fourth Ed., Prentice Hall.

[36] Hansen, L.P. and T.J. Sargent (1993) ,"Seasonality and approximation errors in rational expectations models", *Journal Of Econometrics* 55, 21-55.

[37] Harvey, A.C. (1998) "Long Memory in Stochastic Volatility," in J. Knight and S. Satchell (eds.) *Forecasting Volatility in Financial Markets*, Butterworth-Heinemann, Oxford.

[38] Judge,G., W.E.Griffith, R.C. Hill, H. Lutkepohl and T.-C. Lee (1985) *The theory and Practice of Econometrics*, Second Edition, Jhon Wiley & Sons.

[39] Merton, R. C. (1980), "On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics*, 8, 323-361.

[40] Nelson, D.B. and C.Q. Cao (1992), " Inequality Constraints in the Univariate GARCH Model" *Journal of Business and Economic Statistics*, 10, 229-235.

[41] Sims, C. A. (1974), "Distributed lags", in Intrilligator, M. D. and D. A. Kendrick, *Frontiers of Quantitative Economics II*, North-Holland, Amsterdam.

[42] Sims, C.A. (1993), "Rational expectations modeling with seasonally adjusted data", *Journal Of Econometrics*, 55, 9-19.

[43] Sinko, A. (2003), "Estimation and Specification of Long Memory MIDAS regressions", Department of Economics, UNC, Work in progress.

[44] Stock, J.H. and M.W. Watson (2003), *Introduction to Econometrics*, Addison-Wesley.

[45] Taylor, S. J. and X. Xu (1997), "The incremental volatility information in one million foreign exchange quotations", *Journal of Empirical Finance*, 4, 317-340.

[46] Woerner J. (2002) "Variational sums and power variation: a unifying approach to model selection and estimation in semimartingale models", Discussion Paper, Oxford University.

[47] Wooldridge, J.M. (1999), *Introductory Econometrics: A Modern Approach*, South-Western.

[48] Wood, R.A., T. H. McInish and J. K. Ord (1985) "An Investigation of Transactions Data for NYSE Stocks", *Journal of Finance*, 40, 723-739.

Table 1: Comparisons of MIDAS Models with Daily Regressors

The empirical results are based on a data set of 24 hour FX five-minute intraday DM/US$ quotes covering a ten year period, 1/12/1986 to 30/11/1996. The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Entries to the table represent adjusted in-sample R^2 of one week ($1wk$) through four weeks ($4wks$) prediction horizons. All MIDAS regressions involved Beta distributed lags with only a single parameter estimated (i.e. $\theta_1 = 1$ appearing in equation (1.2)).

	MIDAS with daily lags of regressors				
	$\tilde{Q}^{(m)}_{t,t-1}$	$(r_{t,t-1})^2$	$\|r_{t,t-1}\|$	$[hi - lo]_{t,t-1}$	$\tilde{P}^{(m)}_{t,t-1}$
	k^{max}=50 Days				
1 wk	0.358	0.291	0.336	0.386	0.401
2 wks	0.374	0.276	0.348	0.410	0.427
3 wks	0.311	0.170	0.237	0.331	0.391
4 wks	0.405	0.272	0.341	0.414	0.437
	k^{max}=250 Days				
1 wk	0.360	0.293	0.335	0.387	0.402
2 wks	0.375	0.277	0.346	0.411	0.427
3 wks	0.313	0.171	0.236	0.331	0.391
4 wks	0.405	0.275	0.340	0.413	0.437

Table 2: Estimated weights of models with daily regressors (50 Days)

	μ_H	day 1	day 5	days 6-10	days 11-20	days 21-30	> 30 days
			Realized Volatility (Unconstrained model)				
1 wk	1.111	1.425	0.344	-	-	-	-
2 wks	2.415	2.163	0.569	-	-	-	-
3 wks	3.379	4.430	1.149	-	-	-	-
4 wks	4.529	3.930	0.313	-	-	-	-
			MIDAS with daily lags of $(r_{t,t-1})^2$				
1 wk	1.342	0.293	0.195	0.706	0.578	0.118	0.012
2 wks	2.942	0.484	0.334	1.250	1.110	0.264	0.032
3 wks	4.825	0.344	0.297	1.320	1.887	1.050	0.502
4 wks	6.360	1.008	0.658	2.358	1.866	0.359	0.033
			MIDAS with daily lags of $\tilde{Q}^{(m)}_{t,t-1}$				
1 wk	0.857	0.708	0.296	0.770	0.263	0.010	0.000
2 wks	1.963	1.423	0.536	1.297	0.368	0.010	0.000
3 wks	3.141	2.128	0.815	1.995	0.583	0.017	0.000
4 wks	4.232	4.300	0.845	1.362	0.136	0.000	0.000
			MIDAS with daily lags of $\|r_{t,t-1}\|$				
1 wk	-0.470	0.855	0.488	1.581	0.957	0.112	0.005
2 wks	-0.298	1.472	0.862	2.845	1.810	0.233	0.012
3 wks	0.516	1.414	0.999	3.796	3.526	0.909	0.126
4 wks	0.601	3.136	1.624	4.904	2.469	0.204	0.006
			MIDAS with daily lags of $\tilde{P}^{(m)}_{t,t-1}$				
1 wk	-1.696	0.230	0.029	0.036	0.002	0.000	0.000
2 wks	-2.478	0.467	0.038	0.038	0.001	0.000	0.000
3 wks	-4.070	0.586	0.093	0.133	0.010	0.000	0.000
4 wks	-4.489	0.979	0.065	0.059	0.001	0.000	0.000
			MIDAS with daily lags of $[hi - lo]_{t,t-1}$				
1 wk	-0.938	85.454	31.966	77.096	21.680	0.566	0.003
2 wks	-1.240	154.377	57.661	138.929	38.968	1.012	0.006
3 wks	-1.595	195.838	87.382	237.645	91.050	4.419	0.058
4 wks	-1.584	328.092	106.280	233.149	51.429	0.813	0.002

Table 3: **Estimated weights of models with daily regressors (250 Days)**

	μ_H	day 1	day 5	days 6-10	days 11-20	days 21-30	> 30 days
			Realized Volatility (Unconstrained model)				
1 wk	1.111	1.425	0.344	-	-	-	-
2 wks	2.415	2.163	0.569	-	-	-	-
3 wks	3.379	4.430	1.149	-	-	-	-
4 wks	4.529	3.930	0.313	-	-	-	-
			MIDAS with daily lags of $(r_{t,t-1})^2$				
1 wk	1.298	0.301	0.190	0.678	0.581	0.170	0.063
2 wks	2.843	0.487	0.324	1.198	1.119	0.374	0.168
3 wks	4.678	0.388	0.309	1.304	1.696	0.922	0.969
4 wks	6.142	0.998	0.636	2.288	1.990	0.596	0.228
			MIDAS with daily lags of $\tilde{Q}^{(m)}_{t,t-1}$				
1 wk	0.842	0.728	0.287	0.748	0.293	0.024	0.002
2 wks	1.946	1.461	0.519	1.262	0.419	0.026	0.002
3 wks	3.122	2.289	0.780	1.851	0.576	0.033	0.002
4 wks	4.223	4.352	0.827	1.386	0.182	0.002	0.000
			MIDAS with daily lags of $\|r_{t,t-1}\|$				
1 wk	-0.494	0.881	0.472	1.505	0.970	0.183	0.038
2 wks	-0.339	1.512	0.833	2.704	1.827	0.370	0.084
3 wks	0.388	1.483	0.978	3.603	3.328	1.091	0.477
4 wks	0.479	3.145	1.569	4.765	2.726	0.424	0.070
			MIDAS with daily lags of $\tilde{P}^{(m)}_{t,t-1}$				
1 wk	-1.709	0.230	0.029	0.039	0.003	0.000	0.000
2 wks	-2.505	0.465	0.040	0.044	0.002	0.000	0.000
3 wks	-4.096	0.587	0.093	0.141	0.014	0.000	0.000
4 wks	-4.497	0.983	0.065	0.064	0.002	0.000	0.000
			MIDAS with daily lags of $[hi - lo]_{t,t-1}$				
1 wk	-0.959	86.452	31.060	76.148	25.722	1.676	0.102
2 wks	-1.261	157.574	55.771	135.446	44.701	2.799	0.163
3 wks	-1.627	203.644	84.343	226.512	95.732	9.093	0.840
4 wks	-1.607	331.119	103.221	231.614	62.918	2.815	0.113

Table 4: Comparisons of Other Models

The empirical results are based on a data set of 24 hour FX five-minute intraday DM/US$ quotes covering a ten year period, 1/12/1986 to 30/11/1996. The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Entries to the table represent adjusted in-sample R^2 of one week ($1wk$) through four weeks ($4wks$) prediction horizons for the *unconstrained* regression appearing in equation (1.8) and the MIDAS regression appearing in equation (1.4) involved Beta distributed lags with only a single parameter estimated (i.e. $\theta_1 = 1$ appearing in equation (1.2)).

		Realized Volatlity 30 min.		Realized Volatility 5 min.		MIDAS $(r_{t,t-1})^2$	
k^{max}		1 day	5 days	1 day	5 days	50 days	250 days
	1 wk	0.131	0.196	0.267	0.337	0.291	0.293
	2 wks	0.138	0.239	0.247	0.350	0.276	0.277
	3 wks	0.155	0.241	0.212	0.291	0.170	0.171
	4 wks	0.186	0.337	0.275	0.386	0.272	0.275

Table 5: Comparison of models with high-frequency regressors

The empirical results are based on a data set of 24 hour FX five-minute intraday DM/US$ quotes covering a ten year period, 1/12/1986 to 30/11/1996. The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Using this FX data set we calculate intra-day volatilities for 5- and 30-minute time intervals. Entries to the table represent adjusted in-sample R^2 of one week ($1wk$) through four weeks ($4wks$) prediction horizons. Beta distributed lag MIDAS regressions (appearing in equation (1.2)) are used with parameters θ_1 and θ_2 unrestricted.

	1 day	5 days	1 day	5 days
	30 minutes		5 minutes	
Daily Realized Volatility (Unconstrained model)				
1 wk	0.137	0.196	0.272	0.337
2 wks	0.138	0.239	0.247	0.350
3 wks	0.155	0.241	0.212	0.291
4 wks	0.186	0.337	0.275	0.386
MIDAS with intra-daily lags of $(r_{t,t-1/m})^2$				
1 wk	0.147	0.205	0.274	0.331
2 wks	0.197	0.259	0.302	0.336
3 wks	0.156	0.215	0.283	0.290
4 wks	0.251	0.319	0.323	0.391
MIDAS with intra-daily lags of $\lvert r_{t,t-1/m}\rvert$				
1 wk	0.223	0.279	0.371	0.395
2 wks	0.306	0.330	0.394	0.415
3 wks	0.230	0.320	0.352	0.380
4 wks	0.301	0.362	0.409	0.437

	50 days	250 days	50 days	250 days
	$(r_{t,t-1/m})^2$		$\lvert r_{t,t-1/m}\rvert$	
1 wk	0.357	0.358	0.398	0.399
2 wks	0.372	0.373	0.424	0.425
3 wks	0.312	0.313	0.389	0.389
4 wks	0.398	0.397	0.428	0.428

Table 6: Comparison of models with hourly adjusted HF data

The empirical results are based on a data set of 24 hour FX five-minute intraday DM/US$ quotes covering a ten year period, 1/12/1986 to 30/11/1996. The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Using this FX data set we calculate intra-day volatilities for 5- and 30-minute time intervals. Entries to the table represent adjusted in-sample R^2 of one week ($1wk$) through four weeks ($4wks$) prediction horizons. Beta distributed lag MIDAS regressions (appearing in equation (1.2)) are used with parameters θ_1 and θ_2 unrestricted. All returns are adjusted for hourly seasonal variation.

	1 day	5 days	1 day	5 days
	30 minutes		5 minutes	
	Hourly adjusted $(r_{t,t-1/m})^2$			
1 wk	0.095	0.186	0.233	0.315
2 wks	0.057	0.227	0.288	0.315
3 wks	0.124	0.196	0.243	0.245
4 wks	0.210	0.310	0.308	0.353
	Hourly adjusted $\|r_{t,t-1/m}\|$			
1 wks	0.193	0.237	0.347	0.366
2 wks	0.268	0.280	0.362	0.382
3 wks	0.189	0.273	0.322	0.328
4 wks	0.254	0.323	0.374	0.373

Table 7: Comparison fractional differencing models

The empirical results are based on a data set of 24 hour FX five-minute intraday DM/US$ quotes covering a ten year period, 1/12/1986 to 30/11/1996. The final sample includes 705,024 five-minute returns reflecting 2,448 trading days. Using this FX data set we calculate intra-day volatilities for 5- and 30-minute time intervals. Entries to the table represent adjusted in-sample R^2 of one week ($1wk$) through six weeks ($4wks$) prediction horizons. Beta distributed lag MIDAS regressions (appearing in equation (1.2)) are used with parameters θ_1 and θ_2 unrestricted..

	1 day	5 days	1 day	5 days	1 day	5 days
	Unc. $\tilde{Q}^{(m)}_{t,t-1}$		FI MIDAS $(r_{t,t-1/m})^2$		FI MIDAS $\|r_{t,t-1/m}\|$	
1 wk	0.339	0.337	0.370	0.354	0.377	0.372
2 wks	0.323	0.314	0.375	0.372	0.408	0.400
3 wks	0.264	0.263	0.391	0.343	0.405	0.376
4 wks	0.228	0.260	0.308	0.332	0.373	0.362

	50 days	250 days	50 days	250 days
	FI MIDAS $\tilde{Q}^{(m)}_{t,t-1}$		FI MIDAS $\tilde{P}^{(m)}_{t,t-1}$	
1 wk	0.333	0.332	0.367	0.366
2 wks	0.370	0.320	0.404	0.401
3 wks	0.348	0.267	0.381	0.270
4 wks	0.333	0.333	0.363	0.242
	FI MIDAS $[hi - lo]_{t,t-1}$		FI MIDAS $(r_{t,t-1/m})^2$	
1 wk	0.368	0.368	0.348	0.331
2 wks	0.398	0.398	0.370	0.319
3 wks	0.357	0.272	0.346	0.265
4 wks	0.360	0.233	0.327	0.233

FI MIDAS $|r_{t,t-1/m}|$

	50 days	250 days
1 wk	0.365	0.364
2 wks	0.402	0.401
3 wks	0.381	0.264
4 wks	0.356	0.237

Table 8: Parameter Estimates of selected models with 50 lags of daily data

	μ_H	$s.e.$	ϕ_H	$s.e.$	θ_2	$s.e.$
			MIDAS with daily lags of $\tilde{Q}^{(m)}_{t,t-1}$			
1 wks	0.857	0.065	3.453	0.101	11.240	1.203
4 wks	4.232	0.130	12.767	0.208	20.110	0.775
			MIDAS with daily lags of $\tilde{P}^{(m)}_{t,t-1}$			
1 wks	-1.696	0.129	0.567	0.016	25.463	2.755
4 wks	-4.489	0.253	2.000	0.031	32.939	1.672
			FI MIDAS with daily lags of $\tilde{Q}^{(m)}_{t,t-1}$			
1 wks	2.274	0.297	0.703	0.273	0.975	0.059
4 wks	6.299	0.319	6.999	0.234	23.394	1.264
			FI MIDAS with daily lags of $\tilde{P}^{(m)}_{t,t-1}$			
1 wks	1.422	0.277	0.174	0.014	300.000	1137.126
4 wks	0.545	0.419	1.244	0.038	38.295	2.364

Table 9: Parameter Estimates of selected models with 5 days of intra-daily data

	μ_H	$s.e.$	ϕ_H	$s.e.$	θ_1	$s.e.$	θ_2	$s.e.$
				MIDAS with intra-daily lags of $(r_{t,t-1/m})^2$				
1 wks	1.139	0.054	844.580	14.564	0.912	0.009	1.769	0.144
4 wks	4.320	0.131	3540.697	58.275	1.455	0.074	2.463	0.146
				MIDAS with intra-daily lags of $\|r_{t,t-1/m}\|$				
1 wks	-1.638	0.121	160.360	4.239	0.907	0.008	1.939	0.208
4 wks	-4.145	0.249	552.992	8.700	1.788	0.096	5.916	0.451
				FI MIDAS with intra-daily lags of $(r_{t,t-1/m})^2$				
1 wks	2.384	0.252	189.411	24.112	0.979	0.066	7.941	2.133
4 wks	6.237	0.323	1965.182	64.858	1.585	0.166	2.280	0.241
				FI MIDAS with intra-daily lags of $\|r_{t,t-1/m}\|$				
1 wks	1.296	0.283	54.417	5.270	0.933	0.018	6.781	2.218
4 wks	0.490	0.403	353.965	10.233	1.710	0.124	5.336	0.512

Figure 1: MIDAS Weights

This figure plots the weights (normalized to sum up to one) obtain from an OLS regression for a four week ($4wks$) prediction horizon for (1) a regression model based on daily squared returns (2)a regression involving past daily realized volatility, denoted daily QV (3) daily realized, denoted PV, (4) daily absolute returns, denoted ABS and finally (5) the daily range denoted R.

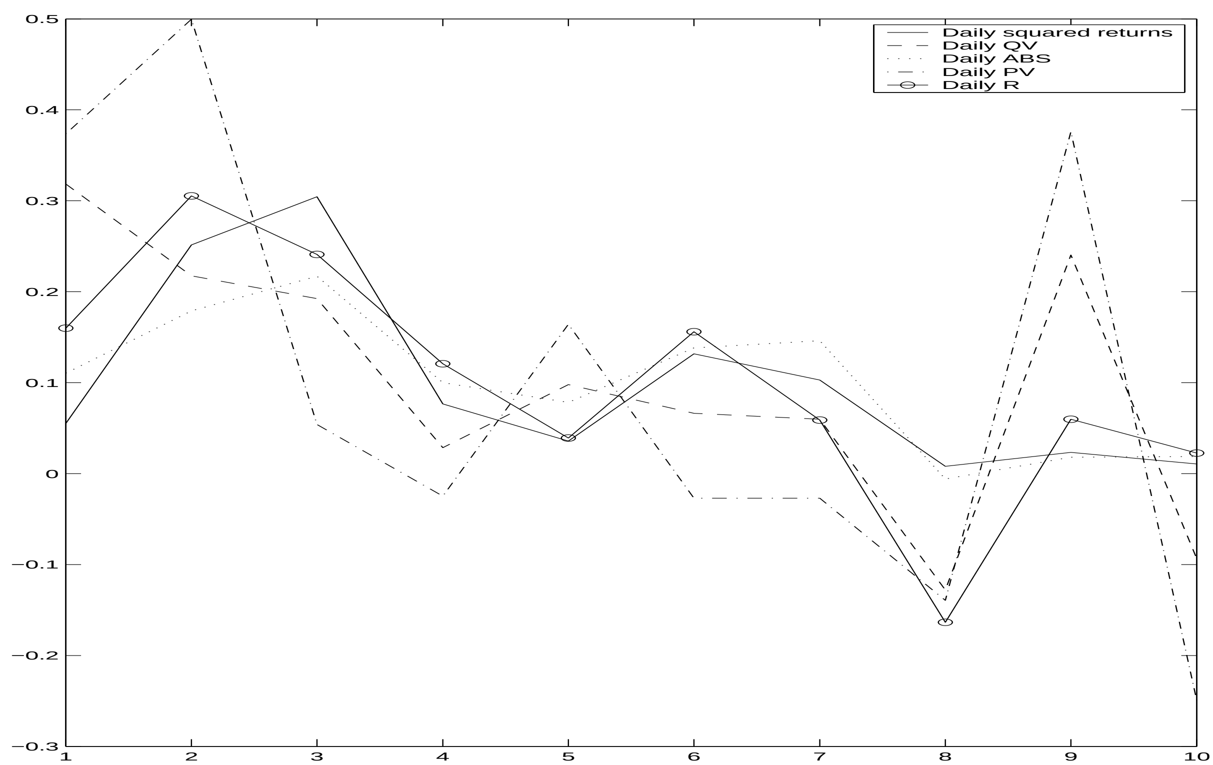

Figure 2: MIDAS Weights One Week Horizon and Daily regressors

This figure plots the weights that the MIDAS estimators of one week ($1wk$) prediction horizon for (1) a MIDAS model based on daily squared returns appearing in regression (1.9)(2) MIDAS regression involving past daily realized volatility appearing in regression (1.1) denoted QV (3) daily realized power to run regression (1.7) denoted PV, (4) daily absolute returns to run regression (1.5) denoted ABS and finally (5) the daily range for regression (1.6) denoted R.

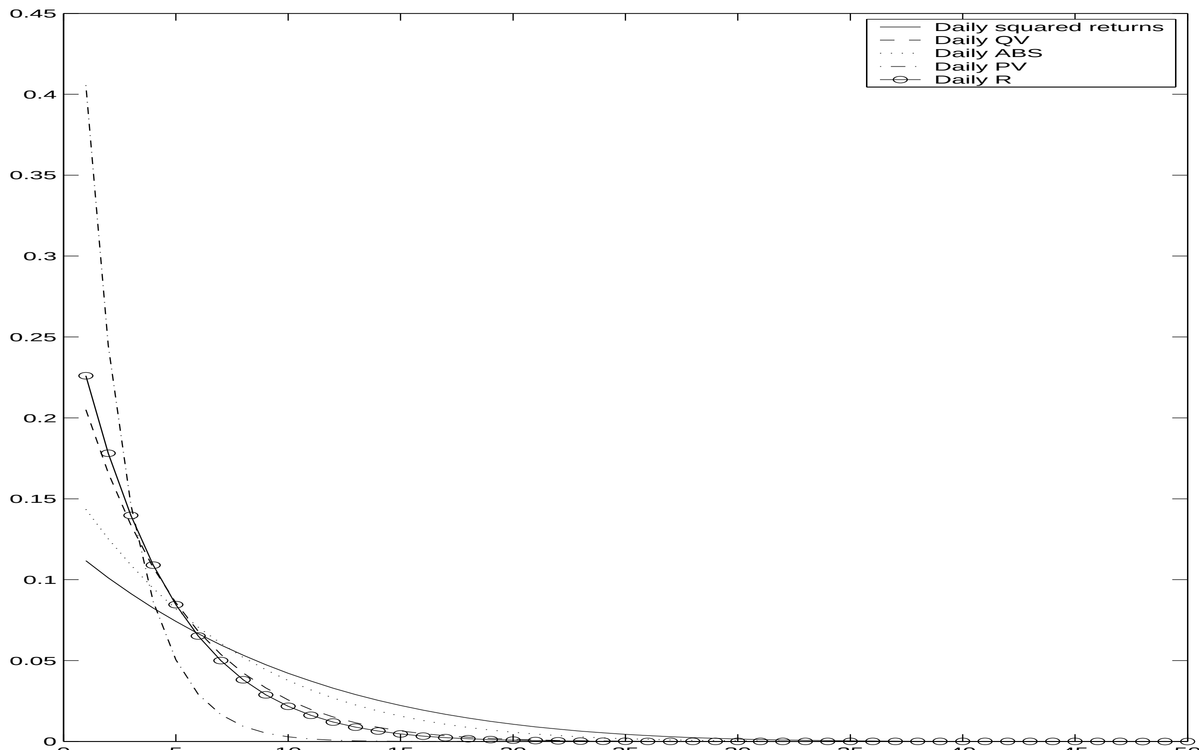

Figure 3: MIDAS Weights Four Weeks Horizon and Daily regressors

This figure plots the weights that the MIDAS estimators of four weeks ($4wks$) prediction horizon for (1) a MIDAS model based on daily squared returns appearing in regression (1.9)(2) MIDAS regression involving past daily realized volatility appearing in regression (1.1) denoted QV (3) daily realized power to run regression (1.7) denoted PV, (4) daily absolute returns to run regression (1.5) denoted ABS and finally (5) the daily range for regression (1.6) denoted R.

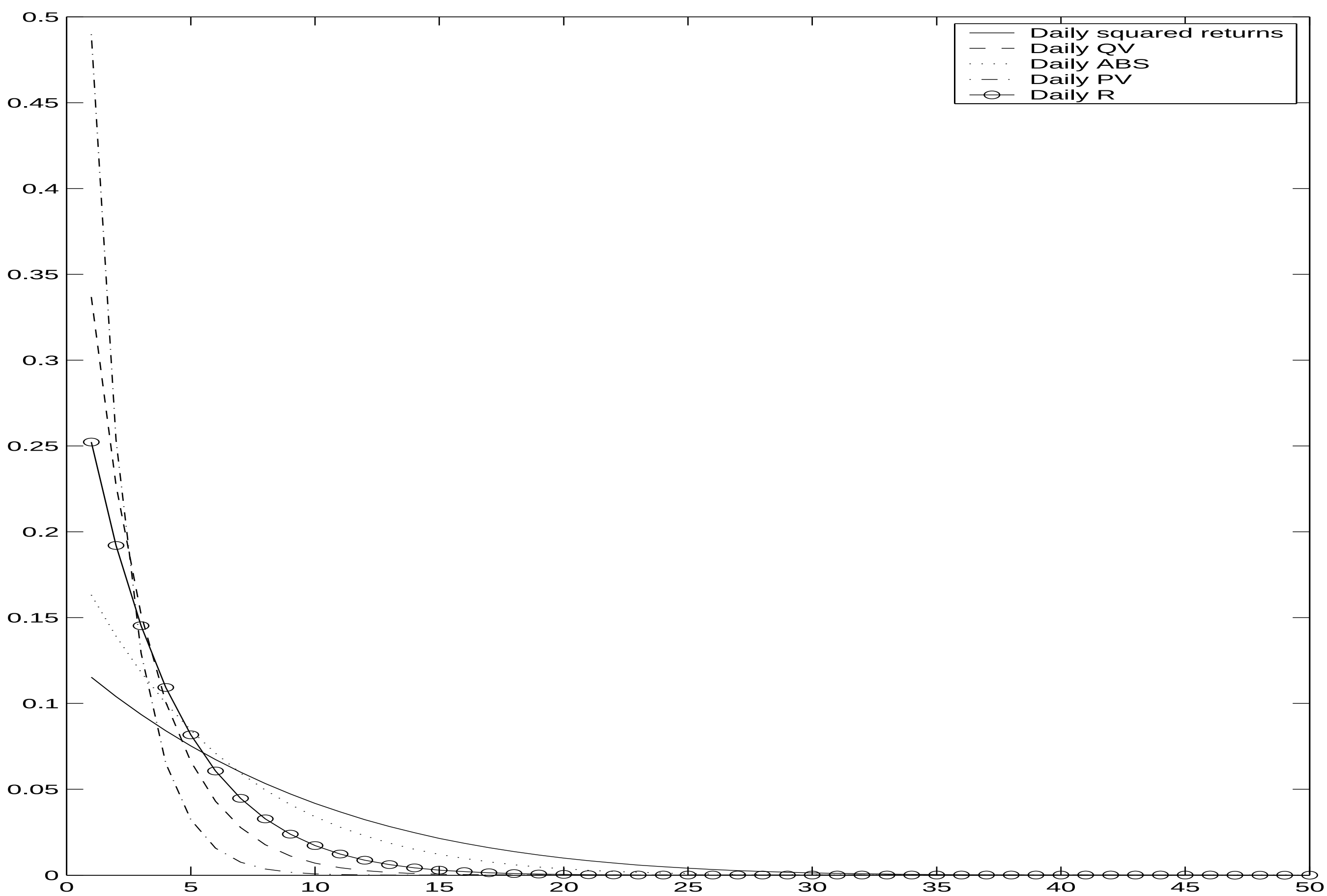

Figure 4: MIDAS Weights One Week Horizon and Intra-daily regressors

This figure plots the weights that the MIDAS estimators of one week ($1wk$) prediction horizon for (1) a MIDAS model based on high-frequency intra-daily squared returns appearing in regression and (2) high-frequency absolute returns.

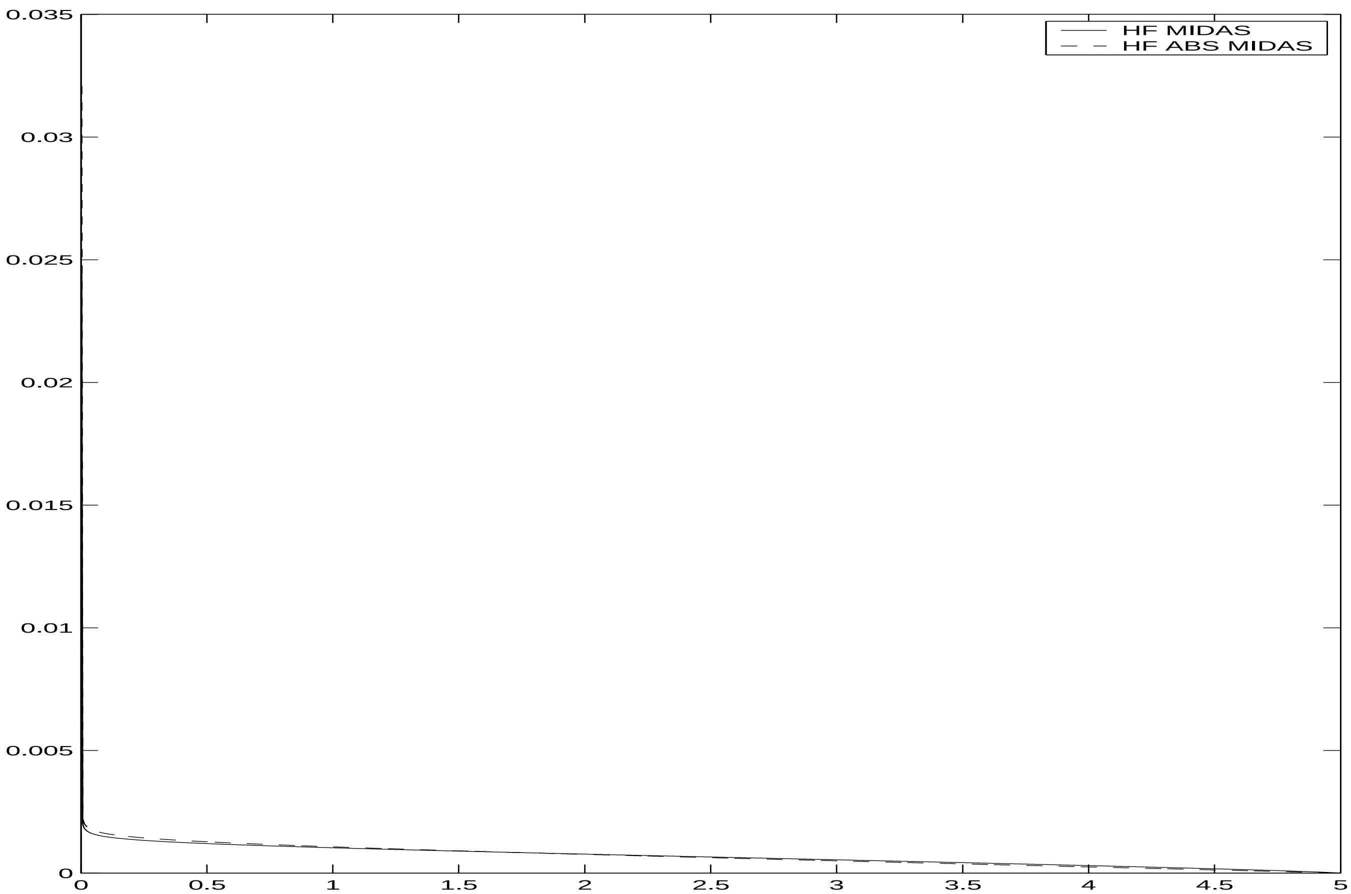

Figure 5: MIDAS Weights Four Weeks Horizon and Daily regressors

This figure plots the weights that the MIDAS estimators of four weeks ($4wks$) prediction horizon for (1) a MIDAS model based on high-frequency intra-daily squared returns appearing in regression and (2) high-frequency absolute returns.